

## Rethinking the use of Likert scale: tradition or technical choice?

## Repensando o uso da escala Likert: tradição ou escolha técnica?

**Rafael Lucian**

Faculdade Boa Viagem | DeVry, Recife, PE, Brazil

### ABSTRACT

The decision-making process of the executive is directly conditioned the quality of information they receive, which ultimately empirical approaches, is the result of the measurement instrument 's ability to use valid, accurate answers on the investigation public. In this sense, much of the information obtained by survey and made available to executives are limited Likert scale capabilities to capture the intended construct. Therefore, in order to enhance a discussion that is of great practical and theoretical contribution to the administration, this theoretical essay propose to investigate whether there is some kind of an attitude scale of measurement that is most appropriate for use in management of the traditional model of Likert five points. To do so is made a literature review in depth and in the end, a new scale of measurement more accurate and balanced is proposed.

**KEYWORDS:** Attitude of measurement, Likert scale, opinion research.

### RESUMO

O processo decisório dos executivos é diretamente condicionado à qualidade da informação que recebem que em abordagens empíricas é fruto da capacidade do instrumento de mensuração utilizado para respostas válidas e precisas sobre o público investigado. Nesse sentido, grande parte das informações obtidas por meio de pesquisas de opinião e disponibilizadas aos executivos estão limitadas à capacidade da escala Likert de captar o construto entendido. Com o objetivo de contribuir com os profissionais de Administração, este ensaio teórico se propõe a investigar se existe algum tipo de escala de mensuração de atitude que seja mais indicada para o uso em Administração do que o modelo tradicional de Likert de 5 pontos. Para tanto, foi feita uma profunda revisão da literatura e, ao final, uma nova escala de mensuração mais precisa e balanceada é proposta.

**PALAVRAS-CHAVE:** Mensuração de atitude, escala Likert, pesquisa de opinião.

Submission: 26 October 2015

Approval: 21 January 2016

### **Rafael Lucian**

Ph.D. in Business Administration from Universidade Federal de Pernambuco (UFPE). Center for Research Support Coordinator at Boa Viagem College (FBV) | DeVry.

(CEP 51200-060 – Recife, PE, Brazil).

E-mail: rluccian@fbv.edu.br

Address: Campus Imbiribeira, Rua Jean Émile Favre, 422 - Imbiribeira - 51200-060, Recife, PE, Brazil.

## 1 INTRODUCTION

Today's society has several contours of modernity and intensive use of information and communication technology, but there is still a growing area for concern for the individual, their nature and their behavior (PETTER; STRAUB, 2012).

This effort to understand human behavior, with forays into psychology, in order to perceive and capture the individual's attitudes is a colossal task that has its basis in Psychometrics. Business administration authors such as Lawrence and Lorsch (1967), Parasuraman et al. (1988), Venkatesh et al. (2003) and Mittelman (2014) made use of psychometric principles in the service of organizational studies.

Attitude is characteristic of the people refers to the set of beliefs about something and their answer about it (THURSTONE, 1928). Measuring attitude is important because this knowledge is useful in understanding the behavior of people according to their decision process.

The first approaches dating from the nineteenth century, but studies have migrated and have intensified in Social Sciences between 1920 and 1930. The early studies meeting this goal through individualized psychological tests, which is a method that required many resources and had low ability to reach large groups (THURSTONE, 1928).

On the other hand, due to the great acceptance that the attitude's research obtained at the Academy of Psychology, there was scholars' efforts to develop methods that enable its application in groups, among them, stood out the proposition Likert (1932) which stated that measuring the attitude groups was possible using multi-item scale.

This new form of measurement scale, called multi-item measure is a scientific instrument of observation and measurement of idealized social phenomena in order to measure attitudes through opinions objectively (LIKERT, 1932). This proposal received wide recognition by the academy and it applies to the present day. Opinion scales also name the scales with this purpose.

From this perspective, although the Likert scale is widely accepted and stable, there are questions about it which so far have not been resolved, such as how to analyze the data and ordinal variables that allow only non-parametric tests, as claimed by Rasmussen et al. (1989), and Jöreskog Sörbom (1996) and Schriesheim and Castro (1996).

The limited data from Likert scale is so severely that some authors as Ockert (2005), for example, states that the data resulting from the scale in question allow only relational conjectures. It is said that the data from a Likert scale, only allow ordering of individuals through the favorability of their attitude towards a particular object, but do not reveal a basis to say how much an individual is more favorable than another, or to measure the amount of change in a certain experience (SELLTIZ et al, 1975;. SANCHES; MEIRELES; SORDI, 2011).

Among the other possible existing gaps in Likert scale is the issue of scale gradation, according to Preston and Coleman (2000), it has not been resolved even after discussions decades. Emphasizing the presence of this failure, Vieira and Dalmoro (2008) state that is very common surveys showing in his methodological procedure the expression that used a Likert scale of  $n$  points, but there is rare appearance justifications for the choice of the value of  $n$ .

Thus, facing this clear research opportunity, this theoretical essay proposes a general objective to investigate if there is any kind of attitude measurement scale that is more suitable for use in management than the traditional model of Likert five points.

For this purpose, are presented below an introduction to measurement, an in depth review of the original proposition Likert (1932), the main criticism of the multi-item model five points and at the end, an alternative that reduces the criticisms known to benefit an instrument with greater power measurement.

## **2 MEASUREMENT OF SCALES: A BRIEF EXPLANATION**

Architecture scale function is to allow the floor plan of a particular space can be made and, even without the objects present, there is the ability to know the ideal location of each item. In the medical application of measurements for scale it has a key role in the diagnosis, because their basic information is at times more accurate than the free description of the patients (BRACHER, 2008).

An example of this is the Graded Chronic Pain Scale (GCPS), which was developed to provide a simple and brief method to grade the severity of chronic or recurrent pain (VON KORFF; MOORE, 2001). In this case, the construct is the pain and the measurement is made by the patient's expression in a staggered visual scale from zero to ten for each of the issues.

Scales used in product and process measurement constitute a particular type, known as parametric scales, as it characterizes the absoluteness of numbers. It turns out that this kind of scale does not exist in Psychology or in its branches such as the measurement of attitude, where even the Quotient test of intelligence (IQ) is heavily criticized for not achieving this parametric feature (PATTO, 1997).

Thus, a century after Metre Convention in 1799, Galton (1880) dug the term nonparametric scale for new scales which were devoted to measure the psychological rather than the physical world.

The main characteristic of non-parametric scales is their inability to provide absolute results because each scale provides a result that should be compared under very strict conditions so that there is construction of knowledge and this is part of the challenge of metrics that purport to measure the psychological known as psychometric.

The Psychometrics is the measurement of psychological variables in order to measure and compare individuals and groups (MICHELL, 1997). To achieve this goal, Pasquali (2003) emphasized the importance of scale arising from the social psychology, such as the Likert scale.

Netemeyer, Bearden and Sharma (2003) define psychometrics such as the measurement of psychological attributes that are not directly observable, but which may have its presence evidenced by a set of behaviors or opinions.

In Business Administration field one of the most traditional ways of using scales is psychometrics, it is of interest to managers observe and understand the behavior of individuals and groups through his attitude. This format, however, admits certain rigors for its use and analysis, which can only be discussed after a brief review of the types of scales and their characteristics.

### **2.1 TYPES OF SCALE: EARLY STUDIES**

The types of scales were originally proposed by Stevens (1946) and they are nominal, ordinal, interval and ratio. This classification was proposed to be noted that the previous classification of intensive and extensive was no longer sufficient to explain the diversity of scales that emerged in the field of psychology. The intention of the author, fundamental to this section, was to create a set of measurement rules linked to each type of scale, which would explain the applications of each.

The nominal scale, where there are no numerical relationship between the degrees represents the use of labels that only serve as titles for the scale points and could be replaced without loss by letters, names or images. The only statistical allowed for this group is the relative frequency of cases. This type is a fairly simple way to scale and of course many do not give it due importance. It has only a single rule that is not designating the same numbers for different classes or different numbers for the same classes. An example of the nominal scale is a list of students' enrollments.

The ordinal scale that preserves the ordering of numbers, but not of their ranges, arose from the operation of the gradual escalation, with the characteristic to preserve the order of the points of invariable form. This kind of scale does not allow the calculation of mean and standard deviation and its use only reveals the relative position of the object on a gradual scale.

Stevens (1946) justifies the use of mean and standard deviation are reckless, because scale intervals are not equal in size. It is, however, possible Percentile perform computations beyond count cases and the scale position knowledge. An example of an ordinal scale is the allocation of positions in a ranking of the best bands of the decade.

The use of interval scale in which there is knowledge of the intervals, but there is no definition of absolute zero, is the form most commonly required for quantitative research. Almost all statistics are applicable to this form of measurement, but it provides you with certain limitation is the inability to determine the true zero point, which is determined by convenience or convention. Examples of this type are the temperature ranges Celsius and Fahrenheit, they attach both a zero value to some arbitrary point do not match.

The reason scales, which have absolute zero, are based in the conversion of ownership among different scales measuring the same object, are the most commonly encountered in physics and are the ones with equality, gradual spatial and equal intervals. The absolute zero on the scale is always present, even when abstract.

It is believed that the ratio scale has real meaning for the colloquial use of the word scale because, from day to day, all measures (food, liters, meters, velocity) are examples of this type of scale, although they still can be divided between basic and derived. The first correspond to the basic scales of the universe, such as length and weight, while the derivatives are obtained in fundamental function as the division between the two scales (km/h), but by no means lose their ratio scale properties.

A valuable warning is that the choice of the scale is a function of measurement of capacity used and it does not necessarily imply a quality relationship. The measurement is as accurate as the instrument's ability to capture the empirical fragments. So any scale, sensory or physical, is always subject to bias, low precision and other restrictions.

Although Stevens proposition (1946) is subsequent to the seminal text Likert (1932), their classification is valid for all existing scales, including the psychometric, and therefore standard language between the books and statistical software.

### **3 ORIGINAL LIKERT'S PROPOSAL**

Likert (1932) assumed that attitude could not be measured from a single view, as intended Thurstone (1928), it would actually be the result of a set of opinions. The set of actions that a person can have varies according to the set of stimuli that it can receive. In this sense, capturing the attitude of a person by means of a single view is impossible, according Thurstone (1928), since there is a scale with defined parameters, since the possibilities are endless and attitudes only mark a point on the continuum is impractical.

Thus, the measurement of attitude is, in essence, indirect. Occurs from the opinions, and there are of course a fluctuation of the answers, which can be minimized by considering a set of assertions rather than just a drive as a degree of the scale.

In Likert (1932) research on internationalism, racism, economics, politics and religion, he rigorously followed in the footsteps of Thurstone (1928). Opinions were collected from interviews with experts, statements in newspapers, trade magazines and some books. The statements were issued in order to fit with the basic requirements of simplicity, clarity and brevity. No exceptions, the issues presented should make a value judgment and not facts. This was achieved with the use of statements which referred to future possibilities eliminating a previous judgment of the people and especially the environment. The test was conducted with 2000 respondents, of which only 650 were considered valid.

In spite of previous authors, Likert (1932) proposed measurement by a multi-item scale. As a requirement of their investigation, they decided to use four types of scales:

- Three-point scale;
- Multiple choice Scale with five alternatives;
- Scale 5 points with affirmative and degrees of agreement (most famously);
- Scale 5 points, exactly the same as above, but using news reports.

As this was an innovative use, forms of analysis were also equally innovative. Before this seminal scene, Likert (1932) decided to choose two approaches. In his text he noticed his awareness of its statistical limitations and perhaps why some uncertainty even with the best analytical approach (though obvious, it is necessary to emphasize that was not to use any kind of statistical software at the time and, for any mathematical calculation, it would be necessary to master the algorithms and their properties).

The analysis forms seminal study suggested that, for use in scales were the sum of the sigma points and the distance (distance between the points and the mean). The average distances represented the attitude of the respondent that question and how the attitude is defined by a set of statements and not one that would count was the average of the averages for each cluster. In the analysis by sum of points was considered the value of the alternative selected by the respondent in each statement and at the end, the attitude would be revealed by the sum of these points and not the average.

Some additional methodological observations were recorded in the article and provided the basis for the strong influence in the world of attitude measurement research: the result of the analysis by the sum or average is similar and does not seem to interfere with the final interpretation, being a criterion of choice researcher. The number of alternatives (3, 5 or 7) according Likert (1932) does not interfere with the result.

#### **4 ANALYSIS AND CRITICAL TO VARIOUS MEASUREMENT SCALES OF PROPOSALS**

One of the most frequently critical point about the measurement by scales is that it is not possible, by numbers, measure a complex personality. This is true both for investigations by ramps and for any other form of existing measurement, since no method is able to extract trust with the complexity of human personality.

In fact, measure is to allocate an object somewhere in an abstract continuous (THURSTONE, 1928) and the measurement of any object only describes the attribute (construct) measured. This is a universal feature of the measurements. Thus, the attitude measurement only measures obviously the

attitude. In addition, they can only be measured variables, which can be classified in terms of being higher, or lower than with ample linear profile.

The proposition Thurstone (1928) was to measure attitude by the individual even without accuracy of the answer to the question. He defended himself this aspect arguing that what is measuring is the attitude that the person wants to demonstrate and it shall be in accordance with their actions, it is more likely to be coincidence that contradiction between the answers and attitudes. That is, if the person has such an attitude in his heart, but reveals the company a distinct behavior, the interests of social search is on their explicit attitude and not about what goes on in his head.

At first, the measurement of attitude scales was built on nonparametric statistics. The idea discussed in this section is related to some of the points raised by Galton (1880) that is not carefully observed by the vast majority of contemporary studies in Business Administration.

In fact, a scale such as Allport and Hartman (1925) or Likert (1932) is not metric. At any rate, there will always be a space without values between two statements. It makes a lot of difference from the statistics available for analysis, because there is no external instrument that can be used as a parameter; therefore, any comparison between individuals or groups by these measures would only be possible in relative terms within the same sample.

In order to reduce this problem it is assumed all measured by Likert scale would be about within an arbitrary range (or non-parametric) and thus some statistics are more favorable than others to the scope of measurement. However, even then it is not possible to eliminate this limitation of attitude scales this does not invalidate it.

In detail, if an individual agrees (level 4) with a statement, you can compare it with another that indicated a value of 1 and draw conclusions for. However, if the difference is between neighboring houses, for example, 4 and 5 (strongly agree and agree), there is less confidence in analysis.

But there is, in fact, a recklessness that should be highlighted, since the scale is an instrument that is used to measure aspects of people and not the abstract unity that compose them. In studies of Thurstone (1931) was possible to observe that concern.

As a current example of this theoretical and practical recklessness, cite the usual understanding enshrined in the study of Parasuraman et al. (1988), in which there is belief that SERVQUAL measures the quality of service, when she, in fact, is limited to measure the attitude of people about it. In practice, management actions should aim at practices that would improve the subjective perception of the services or instead of actually believing that parameter quality targets in the services otherwise overwhelm its audience.

According to original theory of Allport and Hartman (1925) and Thurstone (1931), the SERVQUAL should be a graduate continuous in order to place people according to their attitude to the theme of quality of service and, as such, exclusively would measure the attitude of the people and not the quality of service, as already mentioned.

Perhaps the main contribution of Likert (1932) has been the unification of two vectors before measured in independent way, the direction and intensity. The pre-Likert authors used on scale to check respondent's attitude, and another instrument to grade the intensity of such construct.

At the time Likert proposed its scale, it had been already spent half a century of criticism that had condemned the use of average for non-parametric analysis scales. However, his successors did not use the median as suggested, as in a range of only 5 points would be many respondents with the same end result and would reduce, by far, the accuracy of the findings.

It is recalled that the proposal to change the average statistic for the median as the central measure of non-parametric scales is imperative and, given the fact that in an incomplete scale, where only some values are real, any point between the values (intervals) does not exist. For example, if 1 is strongly disagree and 2 disagree, 1.4 does not have any meaning. There are those who think that 1.4 is closer to 1 than 2, but it should be remembered that in a non-metric scale, the intervals are not defined and need not follow any known mathematical criterion, here included the set of real numbers. Thus, even if 1.3 appear closer to 1 than 2, i.e. an inference no mathematical basis, as it is not known if the distance between 3 and 4 is the same as that between 4 and 5. This risk was alerted on the use of non-parametric scales since the nineteenth century.

Likert (1932), however, using a more complex statistical incorporated the statistical analysis, descriptive techniques as mean, standard deviation and sigma distance of the points from the average, even suggests the possibility to use the result of the scales and input for correlations.

The point in question is that, mathematically, Likert (1932) equals the answers of the respondents according to one of the 5-point scale. Moreover, it is impossible to say that the consent of a person in relation to any claim is the same as another respondent. However, this inconsistent interpretation only becomes dangerous when it is enhanced and the embedded error every response is added and turned on average with average errors. In fact, without the intervals are known and knowing that the points are in some degree inaccurate and incomparable intersubjective, the averaging is improper; because there is no way to prove that the weights are right.

Thus, it can be understood that the statistical calculations originally suggested in the application of Likert scale today, after advancing in the study scales are considered unsuitable because both sigma and for the correlation, the problem of range uncertainty exists and prevents the results are reliable, as there had already been alerted by the very creator of the scale.

In fact, in his seminal paper, Likert (1932) suggested that, instead of averaging, the results were treated by means of summation of the points. That measurement study icon took the limitation breaks and said he would not give up his proposition, he believed that, at some future point, this issue would be resolved. However, almost a century later, there is still no certainty about the values of the ranges, although some authors have embarked on this search.

Mager and Kluge (1987) proposed that the two dimensions of meaning and intensity, unified by Likert (1932) were separated again. In this proposal, the respondent first mark if your attitude is positive or negative about something and then would mark a separate scale which the intensity of his attitude, such a proposition is called two-stage scale.

Albaum (1997) stated that there are benefits to using Likert scales two stages over those scales of a stage. The author further states that, by adopting Likert scales two stages, would be avoiding the respondents confuse the dimensions of attitude, forcing in this way, that position themselves individually on each of them. Although traditionally the methodologists as Albaum (1997) do not enter into the issue of data analysis, it is suspected that the result in practice is not as divergent as well, between the scales of one or two stages, as the main limitation of study is the uncertainty interval.

Authors like Kunin (1998) also criticized the use of words on the labels of the scale points. The interpretation of the text is an additional barrier to communication. It was suggested, so that the labels of the scale and replaced by symbols were removed, forming a graphic scale. Figures chosen by Kunin (1998) were designed faces, in which the curvatures of the mouth represent a higher or lower level of happiness.

Although conceptually there is a contribution, Kunin scale (1998) does not change the variable type originated by it, which is the same as the Likert scale. In despite of being discussed in the Management segment as if it were an interval or until same ratio in the extreme cases, for example, in the structural equation modeling does not have this feature, because it has not known ranges. Thus reinforces the warning of the very creator of the scale, that is, the interval question as the main limitation.

Incidentally, if the rigor of Stevens (1946) is applied, the Likert scale (1932) and its variations cannot, neither be classified as ordinals, since it is not possible to say that two scale points do not represent the same thing or two different things that are not represented by the same point. This limitation occurs because you cannot be sure how respondents actually interpret the categories of scale. Although common sense assume that is not likely to ‘completely agree’ an individual is, in fact, equal to the simple ‘agree’ otherwise, it does not, under any circumstances, the scale in question is classified as interval type.

Thus, the ideal rating scale for the study and its variations is nominal since the numbers represent her name (for example, 3 represents not agree or disagree). If so treated and analyzed, any criticism on the lack of rigor is applied (even because the nominal scale is the lowest freedom analysis). If treated as an ordinal, the researcher should point which guarantees support this classification, even in disagreement with Stevens (1946). Important, however, is to ratify that there is no theoretical support in the construction of this scale process to justify it as a reason or interval.

The 5 point scale characteristic and, consequently, the ranges of Likert scale (1932) are the result of attempting to measure the intensity of action. The idea that two people can have the same attitude, but at different intensities is accepted, although its measurement does not appear to be sufficiently precise. In fact, there is no way to make sure that people who mark the same point have the same intensity of attitude: in fact it is much more plausible that do not have and that distance in extreme cases can even reverse the order of the categories, without this is captured by the scale.

The number of categories is arbitrary, and as shown in the actual seminal study, no significant difference ranges between 3 and 5 points. The position that the fact of changing the characteristic of scale, that could be binary, but is multi point characteristic, implies a reduction in data analysis possibilities.

#### **4.1 CRITICAL TO THE ORIGINAL LIKERT’S PROPOSAL**

Although the scale of Likert (1932) is used in many areas, it was originally developed for attitude measurement. The scale was theorized whereas the attitude could not be captured by a single item (then proposed multi-item scale), has developed a way of measuring both the direction and intensity of this action, and this implies the need to use statistics to convert the items into a single construct. It is about these two points that the data analysis should be planned.

The most common method of conversion of items in a single construct is the average statistic, but this practice is not suitable for the type of scale due both restrictions on non-parametric calculations of Galton (1880) and Stevens (1946), as already discussed.

While the construct definition and elaboration of the items is individualized task carried out by the researcher interested in proposing a new scale, the structural definitions of scale are collective, with notable bases in seminal authors. Thus, by adopting a range, statistics possibilities do not require further discussion, because the settings are already known and the criticism should have been considered.



The discussion of data analysis is not part of scales construction process, however, the methodologist is responsible for at least clarify what criteria the analysis must observe. It is not possible discard the importance and need for statistical knowledge of data analysis to create or adopt any scale.

Therefore, it defines the first technical limitation of the Likert scale: the categorization of the variable and the fact that the points of scale represent categories that forbidden the use of most multivariate statistics (STEVENS, 1946; RASMUSSEN et al, 1989; JÖRESKOG; SÖRBOM, 1996; SCHRIESHEIM; CASTRO, 1996). Add to those three criticisms that are subject of debate at the Academy: the neutral point, the labels of the scale and the elements of gradation.

#### **4.1.1 THE NEUTRAL POINT ISSUE**

Discussions on aspects of Likert scale has intensified and one of the most exploited points was the importance of the neutral point. Kamorita (1963), one of the main theorists of this aspect, suggests that it is not possible to clearly define a neutral point on the Likert scale.

This inability to define the neutral point on a graduated scale took Guy and Norvell (1977) to conduct an empirical study in an attempt to draw conclusions about the importance of neutral point on a Likert scale. The authors ascertained that the presence of the neutral point leads respondents to flee the extreme and longer use the midpoints in their responses, indicating that no neutral range can be reliable and accurate. In addition, Garland (1991) interpreted the neutral point serves as a way to override the respondent the question.

Previously, for pragmatic neutral, Peabody (1962) stated that the presence or absence of a neutral category is irrelevant for the validation of the scale, this conclusion also shared by Sjoberg and Nett (1968). Thus, the role of the neutral is to annulling answer or really a neutral attitude.

In fact, attitude has no neutral facet at all, from the time at which the subject becomes aware of the object under examination it will show some positive or negative attitude, which may vary its degree. For example, say that it is indifferent to the company's management can mean positive attitude towards managers, because this is interpretable as a vote of confidence even without explicit support.

Thus, the true function of the neutrals points on the Likert scale is void the issue and not indicate a supposed completely neutral attitude. However, the label 'does not disagree nor agree' not seem to be the best way to say that the point is to set aside the matter in case of inability to reply. For a person who has no knowledge of the plan of positions and salaries of the company its response to the item I am pleased with the plan of positions and salaries of the company should be 'cannot answer' or 'I am not able to respond' rather than 'not disagree nor agree'. In this case, there is a clear difference between a possible neutral attitude and simple ignorance of the measured objective.

#### **4.1.2 LABEL DETERMINATION**

The scales labels format is an ordinary theme in research methodology books, as Malhotra (2006) and Aaker et al. (2007), and is in fact an issue that has received little attention in measurement studies, thus requiring debate on the subject.

Since the beginning, words have been used as labels and numbers, although most of the critical questioning that choice and there has been no success in the alternatives presented by Boyd, Westfall and Stasch (1977) and Alwin (1997): the use of graphic labels. It is recognized, since Reichmann (1964), the visual stimulus is more effective to capture the correct answers to direct questions, as people feel more comfortable reading figures interpreting sentences or marking

numbers. There is also the understanding that the graphics labels are best worked for specific audiences, such as the use of figures for research with children (ALWIN, 1997).

However, this graphical alternative is not accepted as ideal by authors like Chrystal (2008), which stated that the visual language promotes limited term value and the specific respondent's culture can affect your interpretation of the data collection instrument. Given this theoretical impasse, Derham (2011) conducted a series of empirical tests, observing the behavior of three types of Likert scales. The first used only in the word labels, the second used a number series to indicate the degrees of the scale and the third was completely graphic.

The results of the study are shown in Table 1, which versa according to the above study, the most comfortable form the respondent (and therefore more suitable to use) was words as labels. This scale format performed better in six of the seven attributes tested. Apart from the form of labels, the number of degrees on the scale has aroused great interest of the academy, and even today, there is no consensus on its effect on the measurement of attitude.

**TABLE 1**

Label types comparing.

RESPONSES	WORDS	NUMBERS	IMAGES
Preferred	43%	23%	30%
Easiest to answer	40%	21%	29%
Fast to answer	37%	23%	31%
Attractiveness <sup>7</sup>	35%	18%	39%
Best to express feelings	40%	20%	29%
More true answers	45%	23%	23%
Stimulates reflection before answer	27%	25%	36%

Source: Derham (2011, p. 21).

The use of words as labels diverges from the original proposal of Likert once it was used numbers, but between various contemporary applications is already possible to see that, the researchers replaced the numbers with names.

#### 4.1.3 GRADE DESIGN

The Likert scales carries two components: direction and intensity (which can only be measured if there are at least two points for each direction). However, as Cronbach (1951) proposed, it would remain doubts about the effectiveness of measurement of intensity.

Moreover, Bendig (1954) and Komorita (1963) found by empirical investigations that the scale of reliability independent of the number of scale response alternatives. In this latter work, it was found until dichotomous and multiple scales tend to have the same degree of reliability when compared, although nothing can be said about its efficiency.

Analyzing findings like these, Peabody (1962) tested empirically the performance of scales from 2 to 6 points. He found that there is no relevant difference of them to capture intensity information on Likert scales.

Additionally, the issue of measurement efficiency was also observed by Matell and Jacoby (1971) to test ranges from 2 to 19 points in relation to their reliability and validity. These authors concluded that there are no significant differences between them and, more exhaustively, they said that the main component of the Likert scale (1932) is the directional, leaving the intensity a secondary role. The directional component was the most represented important for determining the total power measurement of the graduated scale in the empirical tests performed.

Rodriguez (2005), by meta-analysis concluded that a scale with three response choices is sufficient. This author points out that the effect of the decrease in the number of options to choose shrinks the test proportionally increases efficiency for large amounts of respondents and decreases efficiency for small amounts of respondents. In addition, it shows that the time spent in the questionnaire response is proportional to the total number of alternatives and the use of three items on the scale decreases the time in collecting information.

Wiswanathan, Sudman and Johnson (2004) also show concern for the relationship between the scale and the statistical tests, since the definition of the number of items on the scale will affect the statistical tests to be performed. In fact, the use of a scale that several points may not provide a database of valid for performing statistical inferences, since, according to the sample size may result in a dispersion of respondents, limiting the use of some statistical tests.

Another concern regarding the number of items is on the mismatching of multiple scales (NUNNALLY, 1978). To propose a 5-point scale, the negative intervals tend to be larger than the positive and this unbalanced behavior is independent of the construct that is measuring (TOMAS; OLIVER, 1999).

This effect of intensification of responses and detachment of the ranges is enhanced if the statement referring to the measured item is also written in a negative way (MOTL; DISTEFANO, 2002) and this is one explanation for Barnette (2000) have identified the negative sentences have naturally worse on tests of reliability. In other words, the negative component tends to be overestimated and prevents the correct balancing of scales, or by increasing the interval between the negative items, by the use of positive to negative words.

Rozin explains this asymmetric behavior between positive and negative and Royzman (2001), because of the negative reviews are stronger, intense and faster than positive. Thus, the set of several positive perceptions against only one negative may result in negative attitude, not respecting the arithmetic logic. The suggestion dichotomy of scale, remembering Peabody (1962), is then justified by the asymmetry between positivity and negativity (ANDERSON, 1965).

Thus, for Cacioppo and Berntson (1994), the positive and negative attitudes are independent entities that can coexist, so do not fit the concept of scale. Cacioppo, Berntson and Gardner (1999), then assume that the psychometric measurements are bivariate with bipolar existence of negative and positive without intermediate gradations. Based on these same criteria asymmetry between positive and negative, and by means of empirical tests that measured the intensity of each side of the scale Alexandrov (2010) suggested that the Likert scale pentatonic should not be used for attitude measurement.

## **5 FINAL CONSIDERATIONS AND PROPOSAL OF AN ALTERNATIVE SCALE**

In order to give light to the arguments presented here as critical the original proposition Likert (1932), Chart 1 briefly joins the contributions of the authors already cited on the subject.

**CHART 1**

Main discussion points about Likert scale.

TOPIC	CRITICAL ARGUMENT	AUTHORS
Grade Design	Measuring the direction and intensity of separately attitude gives greater precision to the results.	Mager e Kluge (1987); Albaum (1997); Sanches, Meireles e Sordi (2011); Robertson (2012)
	There is a natural imbalance in the graduated scales, and the negative side is the one with the greatest weight range.	Rozin e Royzman (2001); Motl e DiStefano (2002); Alexandrov (2010)
	There is no definition of what is the ideal number of points in a graduated scale of attitude measurement.	Preston e Coleman (2000); Vieira e Dalmoro (2008)
Neutral Point	The presence or absence of neutral point is irrelevant to the validity of the scale.	Peabody (1962); Sjoberg e Nett (1968)
	The neutral point must be included so that the respondent can undo an issue when you feel the need.	Garland (1991)
Labels	Respondents more easily understand labels graphic scales.	Kunin (1998)
	The use of words on labels is the format that brings the most truthful answers.	Derham (2011)

**Source:** Compilation of the arguments of the works cited in the frame.

## 5.1 PROPOSITION OF A SCALE

Answering the research question of this theoretical essay, it was observed that the literature indicates to bring out a more efficient measurement scale than the original pentatonic Likert. Therefore, it was decided to create a new scale in an attempt to eliminate the limitations observed in that range.

There are three differences between this score and the original Likert scale (1932); use words as labels, the displacement of the neutral point out of the scale and the dichotomy.

With respect to not use numbers as labels followed the recommendation of Derham (2011), which suggests that use labels with words instead of graphics or numbers. This format is more comfortable to the respondent and makes it more prone to understand what is being asked.

The second change was the effective displacement of the neutral point out of the range in order to not provide the continuum of view. Although included and popularized in Likert scale, the neutral point is not mandatory item in graduated intensity scales (PEABODY, 1962; SJÖBERG, NETT, 1968; GUY; NORVELL, 1977). Its use is recommended because it is used in case of not wanting or not having the ability to answer the question; it is therefore a legitimate form of alternative cancellation.

When the neutral point is located in the center of the scale, there is a trend already recognized by the Academy that the answers are there in fact. As noted Guy and Norvell (1977) and Garland (1991) the neutral midpoint attracts the answers for its central location.

Thus, due to the strong connotation that the word void may represent respondents on the scale proposed here we chose to keep indifferent label to the neutral point, with the expectation that is used by the respondent, if this does not have an attitude with respect to any corresponding affirmative.

This choice was because, in fact, the categories of positive and negative attitude are independent and the displacement of the neutral points out that for the respondent, such that its use must be for Response annulment cases when there is no attitude to the statement. If there is any action, regardless of grade, the respondents are forced to position themselves.

Thus, in preparing the scale proposed here, was accepted the suggestion Peabody (1962), Anderson (1965), Cacioppo and Berntson (1994), Cacioppo, Gardner and Berntson (1999) and Alexandrov (2010), to use the positive and negative attitudes as asymmetric positions that should not be measured by scalar quantities; therefore it was decided dichotomized variable.

Dichotomous scales are accepted in the literature and can be used without prejudice to the reliability or effectiveness measurement (BENDIG, 1954; KOMORITA 1961; JACOBY; MATELL, 1971). This format, however, does not allow the intensity to be measured, but it is not believed that this is a disadvantage because, as we have faced before, to theoreticians, the measurement of the intensity proposed by Likert (1932) is impractical due to unbalance between positive and negative attitude (PEABODY, 1962; ALEXANDROV, 2010). It is emphasized that the directional component that is the essence of the attitude measurement (JACOBY; MATELL, 1971) is preserved and in this case enhanced.

## 5.2 PROPOSED SCALE IMPLICATIONS

The scale proposed here has better balance and therefore is more accurate than the original Likert scale (1932). It is believed that future researchers who adopt this scale for measurement will observe more faithfully the results of non-parametric tests, including inter-subjective calculations and between groups.

The new format, dichotomous, still allowing advanced calculations as Item Response Theory (IRT) is used for administration without the need for adaptations and this represents a great advantage, since the mathematical structure of TRI is very complex. Still on analysis of possibilities, the use of binary scale fits perfectly the use of correlations and hypothesis testing nonparametric and even metric calculations as Logistic Regression, depending on other variables involved.

Less well known, but equally important is the fact that the new scale allows the respondent to complete the questionnaire as soon as a simpler scale and visually leaner tired least the individual will tend to have more concentration until the end of the items.

It is expected that such contribution has been adopted and tested by Management Academy so gradually, with the joint effort, the measurements are increasingly accurate and the fruit of these information, the more valuable.

## 6 REFERENCES

AAKER, D. A; KUMAR, V.; DAY, G. S; LAWLEY, M.; STEWART, D. *Marketing research*. Sidney: John Wiley, 2007.

ALBAUM, G. The Likert Scale Revised: An alternative version. *Journal of the Market Research Society*, v. 39, n. 2, p. 1-12, 1997.

ALEXANDROV, A. Characteristics of Single-Item Measures in Likert Scale Format. *The Electronic Journal of Business Research Methods*, v. 8 n. 1, p. 1-12, 2010.

ALLPORT, F. H.; HARTMAN, D. A. The Mensurement and Motivation of Atypical Opinion in a Certain Group. *The American Political Review*, v. 19, n. 4, p.735-760, 1925.

ALWIN, Duane F. Feeling Thermometers Versus 7-Point Scales Which are Better? *Sociological Methods & Research*, v. 25, n. 3, p. 318-340, 1997.

ANDERSON, N. H. Averaging versus Adding as a Stimulus-Combination Rule in Impression Formation. *Journal of Experimental Psychology*, v. 70, n. 4, p. 394-400, 1965.

BARNETTE, J. J. Effects of Stem and Likert Response Option Reversals on Survey Internal Consistency: If You feel the Need, There is a Better Alternative to Using Those Negatively Worded Stems. *Educational and Psychological Measurement*, v. 60, n. 3, p. 361-370, 2000.

BENDIG, A. W. A Factor Analysis of Student Ratings of Psychology Instructors on the Purdue Scale. *Journal of Educational Psychology*, v. 45, n. 7, p. 385, 1954.

BOYD, H. W; WESTFALL, R; STASCH, S. F. *Marketing research*. Text and cases. 4. ed. Illinois: Richard D Irwin Inc., 1977.

BRACHER, E. S. B. *Adaptação e validação da versão em português da escala graduada de dor crônica para o contexto cultural brasileiro*. São Paulo, 2008. Tese (doutorado). Faculdade de Medicina da Universidade de São Paulo.

CACIOPPO, J. T.; BERNTSON, G. G. Relationship between attitudes and evaluative space: A critical review, with emphasis on the separability of positive and negative substrates. *Psychological Bulletin*, v. 115, n. 3, p. 401-423, 1994.

CACIOPPO, J. T; GARDNER, W. L; BERNTSON, G. G. The affect system has parallel and integrative processing components: Form Follows Function. *Journal of Personality and Social Psychology*. v. 76, n. 5, p. 839-855, 1999.

CRONBACH, Lee J. Coefficient alpha and the internal structure of tests. *Psychometrika*, v. 16, n. 3, p. 297-334, 1951.

DERHAM, P. A. J. Using preferred understood or effective scales. How scale presentations effect online survey data collection. *Australasian Journal of Marketing & Social Research*, v. 19, n. 2, 2011.

GALTON, F. Statistics of Mental Imagery. *Mind*, v. 5, n. 19, p. 301-318, 1880.

GARLAND, R. The mid-point on a rating scale: is it desirable? *Marketing Bulletin*, v. 2, p. 66-70, 1991.

GUY, R. F; NORVELL, M. The neutral point on a Likert scale. *The Journal of Psychology*, v. 95, p.199-204, 1977.

INMETRO. *Sistema internacional de unidades – SI*. 8. ed. Rio de Janeiro: 2007.

JACOBY, J.; MATELL, M. Three-point Likert scales are good enough. *Journal of Marketing Research*, v. 8, p. 8, 1971.

JÖRESKOG, K. G.; SÖRBOM, D. *Preliis2: user's reference guide*. *Scientific Software International*. Chicago: 1996.

KOMORITA, S. S. Attitude content, intensity, and the neutral point on a Likert scale. *Journal of Social Psychology*, v. 61, p.327-334, 1963.

KUNIN, T. The construction of a new type of attitude measure. *Personnel Psychology*, v. 51, n. 4, p. 823-824, 1998.

LAWRENCE, P. R.; LORSCH, J. W. *Organization and Environment: Managing differentiation and integration*. Boston: Harvard University, 1967.

- LIKERT, R. A Technique for the Measurement of Attitudes. *Archives of Psychology*, v. 140, p. 1-55, 1932.
- MAGER, J.; KLUGE, A. One-stage versus two-stage attitude measurement: separating the dimensions of direction and intensity. In: *Academy of Marketing Science World Marketing Congress*. Barcelona: 1987.
- MALHOTRA, N. K. *Pesquisa de marketing: uma orientação aplicada*. 4. ed. Porto Alegre: Bookman, 2006.
- MICHELL, J. *Quantitative science and the definition of measurement in psychology*. New York: Brit J Psychol, 1997.
- MITTELMAN, M.; ANDRADE E. B.; CHATTOPADHYAY, A.; BRENDL, C. M. The Offer Framing Effect: Choosing Single versus Bundled Offerings Affects Variety Seeking. *Journal of Consumer Research*, v. 41, n. 4, p. 953-964, Dec. 2014. ISSN: 00935301.
- MOTL, R. W; DISTEFANO, C. Longitudinal Invariance of Selt-Esteem and Method Effects Associated with Negatively Worded Items. *Structural Equation Modeling: A multidisciplinary Journal*, v. 9, n. 4, p. 562-578, 2002.
- NETEMEYER, R. G; BEARDEN, W. O; SHARMA, S. *Scaling Procedures*. Thousand Oaks: Sage Publications, 2003.
- NUNNALLY, J. C. *Psychometric theory*. 2. ed. New York: McGraw-Hill, 1978.
- OCKERT, D. Substantive Scale Verification: A Likert Scale Analysis and Critique of university Student Pedagogical Activity Preferences. *JALT Hokkaido Journal*, v. 9, p. 48-64, 2005.
- PARASURAMAN, A; ZEITHAML, V.; BERRY, L. SERVQUAL: A Multiple-Item Scale for Measuring Consumer Perceptions of Service Quality. *Journal of Retailing*, v. 64, n. 1, p.12-40, 1988.
- PASQUALI, L. A medida psicométrica. In: Pasquali L. *Psicometria: teoria dos testes na psicologia e na educação*. 2 ed. Petrópolis: Editora Vozes, 2003.
- PATTO, M. H. S. Para uma Crítica da Razão Psicométrica. *Psicologia USP*, v. 8, n. 1, p. 47-62, 1997.
- PEABODY, D. Two components in bipolar scales: Direction and extremeness. *Psychologia Revista*, v. 69, p. 65-73, 1962.
- PETTER, S.; RAI, A.; STRAUB, D. The Critical Importance of Construct Measurement Specification: A Response to Aguirre-Urreta and Marakas. *MIS Quarterly*. 36, 1, 147-156, Mar. 2012. ISSN: 02767783.
- PRESTON, C. C.; COLEMAN, A. M. Optimal number of response categories in rating scales: reliability, validity, discriminating power, and respondent preferences. *Acta Psychologica*, n. 104, p. 1-15, 2000.
- RASMUSSEN, J. L. et al. Univariate and Multivariate Groups by Trials Analysis Under Violation of Variance-Covariance and Normality Assumptions. *Multivariate Behavioral Research*, 24, 1, 93, Jan. 1989. ISSN: 00273171.

REICHMANN, William John. *Use and abuse of statistics*, 1964.

ROBERTSON, J. Likert-type Scales, Statistical Methods, and Effect Sizes. *Communications of the ACM*, v. 55, n. 5, 2012.

RODRIGUEZ, M. C. Three options are optimal for multiple-choice items: a meta-analysis of 80 years of research. *Educational Measurement: Issues and Practice*. Summer, 2005.

ROZIN, P; ROYZMAN, E. B. Negativity Bias, Negativity Dominance, and Contagion. *Personality Social Psychology Review*, v. 5, n. 4, p. 296-320, 2001.

SANCHES, C.; MEIRELES, M.; SORDI, J. O. de. Análise qualitativa por meio da lógica paraconsciente: Método de interpretação e síntese de informação obtida por escalas Likert. In: *Anais do III Encontro de Ensino e Pesquisa em Administração e Contabilidade*. João Pessoa: 2011.

SCHRIESHEIM, C.; CASTRO, S. Referent effects in the magnitude estimation scaling of frequency expressions for response anchor sets: an empirical investigation. *Educational and Psychological Measurement*, v. 56, p. 557-569, 1996.

SELLTIZ, C.; JAHODA, M.; DEUTSCH, M.; COOK, S. W. *Métodos de pesquisa nas relações sociais*. São Paulo: EPU, 1975.

SJOBORG, G; NETT, R. *A methodology for social research*. New York: Harper & Row, 1968.

STEVENS, S. S. On the Theory of Scales of Measurement. *Science*, v. 103, n. 2684, p. 677-680, 1946.

THURSTONE, L. L. Attitudes Can Be Measured. *American Journal of Sociology*, v. 33, n. 4, p. 529-554, 1928.

TOMAS, J. M; OLIVER, A. Rosenberg's Self-Esteem Scale: Two Factors or Method Effects. *Structural Equation Modeling*, v. 6, n. 1, p. 84-98, 1999.

VENKATESH, V. et al. User Acceptance of Information Technology: Toward a Unified View. *MIS Quarterly*, v. 27, n. 3, p. 425-478, 2003.

VIEIRA, K. M.; DALMORO, M. Dilemas na construção de escalas tipo Likert: o número de itens e a disposição influenciam nos resultados? *Anais... XXXII ENANPAD*. Rio de Janeiro: 2008.

VON KORFF, Michael; MOORE, James C. Stepped care for back pain: activating approaches for primary care. *Annals of Internal Medicine*, v. 134, n. 9, Part 2, p. 911-917, 2001.

WISWANATHAN, M.; SUDMAN, S.; JOHNSON, M. Maximum versus meaningful discrimination in scale response: implications for validity of measurement of consumer perception about products. *Journal of Business Research*, v.12, n. 57, p. 08-24, 2004.